The State of the Solar Terrestrial Data Environment

Raymond J. Walker

Steven P. Joy

Todd A. King

Paper SH52C-02

Presented at the Fall AGU Meeting

San Francisco, CA

December 6, 2002

How Useful Are Solar Terrestrial Data?

- Modern space instruments provide us with state of the art observations with which to address complex space physics problems.
  - *Observations from older missions help us place current observations in perspective by providing continuity through time.*
- Data that are prepared so that outside scientists can readily use them have a better chance of being useful years from now than data prepared just for the investigation team.
- In this talk we will evaluate the state of space physics data activities from the perspective of scientists who were not involved with the data collection.
  - *Do the data meet the needs of scientists today?*
  - *Do the data meet the requirement to provide a long lasting archive?*

Principles for Successful Science Data Management

[CODMAC Report, Bernstein et al., 1982]

- Scientific Involvement
- Scientific Oversight
  - Peer Review
- Data Availability
  - Timely access determined by the scientific community.
  - Easy to use formats. (Make correlative studies easy.)
  - Appropriate ancillary data
  - Enforce contractual obligations for investigators to place data in the archive.
  - Proper documentation.
- Facilities
- Software
  - Structured, transportable and well documented.
- Scientific data storage
  - Stored in a permanent and retrievable form.
- Data-System Funding
  - Secure funding from mission overruns.

Evaluating Missions Against the CODMAC Principles

1. The data must be accessible.
2. The data must be well documented.
3. The data must be preserved.

1. The data must be accessible

- Are the data available promptly?
- Is it easy for scientists to identify and locate the data needed for a given study?
- Once the data have been located are they readily available to the scientific community?
- Are they available online or on distributable media?
- Are needed calibration and ancillary data available?

Evaluating Missions Against the CODMAC Principles - continued

2. The data must be well documented

- Are the data documented so that knowledgeable scientists who are not instrument experts can use them?
- Does the documentation adhere to recognized standards?
- Is the format of the data described?
- Does the documentation explain how the data were collected, and how they were processed?
- Is data quality including sources of contamination carefully documented?
- Are documents delivered with the data?

Evaluating Missions Against the CODMAC Principles - continued

3. The data must be preserved

- Is there a system in place to assure that the data are not lost?
- Are the data archived on long lasting media?
- Are there copies of the data?
- Is there a program to test and refresh media?

Approach

- We prepared a detailed questionnaire and sent it to the project scientist or project manager for each operating solar terrestrial physics mission.

SOHO (1995)	Geotail (1992)
TRACE (1998)	Cluster (2000)
RHESSI (2002)	Polar (1996)
ACE (1997)	FAST (1996)
Wind (1994)	SAMPEX (1990)
Ulysses (1990)	IMAGE (2000)
Voyager (1977)	TIMED (2001)

Responses received as of January 2, 2003

- Supplemental information was obtained from a recent study of SEC operating missions by Bob McGuire

Data Accessibility

Mission	Validation Period (Months)	Access	Online	Online Request	Percent Available
SOHO	<3	Unlimited	Yes	Yes	75-100
RHESSI	<3	Unlimited[3]	Yes	Yes	75-100
ACE	3-6	Unlimited	Yes	Yes	75-100
Polar	9-12[1]	Unlimited	Yes	Yes	>100
FAST	<3	Unlimited	Yes	Yes	75-100
SAMPEX	<3	Unlimited[2]	Yes[5]	No	75-100
TIMED	6-9	Unlimited	Yes	Yes	25-50
TRACE	<3	Unlimited	Yes	Yes	75-100
Ulysses	<3	Unlimited[4]	Yes	Yes	75-100
IMAGE	<3	Unlimited[3]	Yes	Yes	75-100
Wind	<3	Unlimited	Yes	Yes	75-100

1. Some instruments less
2. Password needed
3. Raw data – project only
4. High resolution data –project only
5. Not WWW accessible

Data Accessibility

- Tested each system by requesting recent data.
  - **SOHO**: Unclear interface, queries took a long time.
  - **RHESSI**: Primitive interface but located and extracted data quickly.
  - **ACE**: Found the data quickly and downloaded it.
  - **Polar**: Simple interface, located data easily and downloaded it.
  - **FAST**: Found the data quickly and downloaded it.
  - **TIMED**: Found the data quickly and downloaded it. Required registration to use.
  - **TRACE**: Nice interface, found data quickly.
  - **Ulysses**: Primitive interface, located data and downloaded it.
  - **IMAGE:** Interface required retyping time interval for each query, found images and downloaded them, had problems with digital data in CDF.
  - **Wind:** Located data through CDAW-Web and downloaded it.

Documentation

	Spacecraft			Instrument			Data Set Overview			Data Processing			Calibration			Quality		
Spacecraft	R	O	V	R	O	V	R	O	V	R	O	V	R	O	V	R	O	V
SOHO	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y	N	-	-	N	-	-
RHESSI	?	?	?	Y	Y	?	N	-	-	Y	Y	?	Y	Y	?	Y	Y	?
ACE	Y	Y	?	Y	Y	?	Y	Y	Y	Y	Y	?	N	-	-	Y	Y	?
Polar	Y	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y	N	Y	Y	N	Y
FAST	Y	Y	Y	Y	Y	?	N	-	-	N	Y	-	N	N	?	N	N	?
SAMPEX	?	?	?	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y
TIMED	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y	Y	Y	Y
TRACE	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?	?
Ulysses	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y
IMAGE	Y	Y	Y	Y	Y	Y	N	-	-	Y	Y	Y	Y	Y	Y	N	-	-
Wind	Y	N	Y	Y	N	Y	N	-	-	Y	N	Y	N	-	-	Y	Y	Y

Key

R-Required

O-Online

V-Verified

Y-Yes

N-No

?-No response received

--NA

Format/File Structure

	Documentation			
Spacecraft	R	O	V	Formats Used
SOHO	Y	Y	Y	**FITS, CDF, SFDU, ASCII**
RHESSI	Y	Y	Y	**FITS, ASCII, Tables**
ACE	Y	Y	?	**HDF, Tables**
Polar	Y	N	Y	**CDF, ASCII, Tables, Other**
FAST	Y	N	Y	**CDF**
SAMPEX	Y	Y	Y	**Other**
TIMED	Y	Y	Y	**Other**
TRACE	?	?	?	**FITS, Other, SFDU**
Ulysses	Y	Y	Y	**CDF, ASCII, Tables, Other**
IMAGE	Y	Y	Y	**CDF, UDF, images**
Wind	Y	Y	Y	**ASCII, CDF**

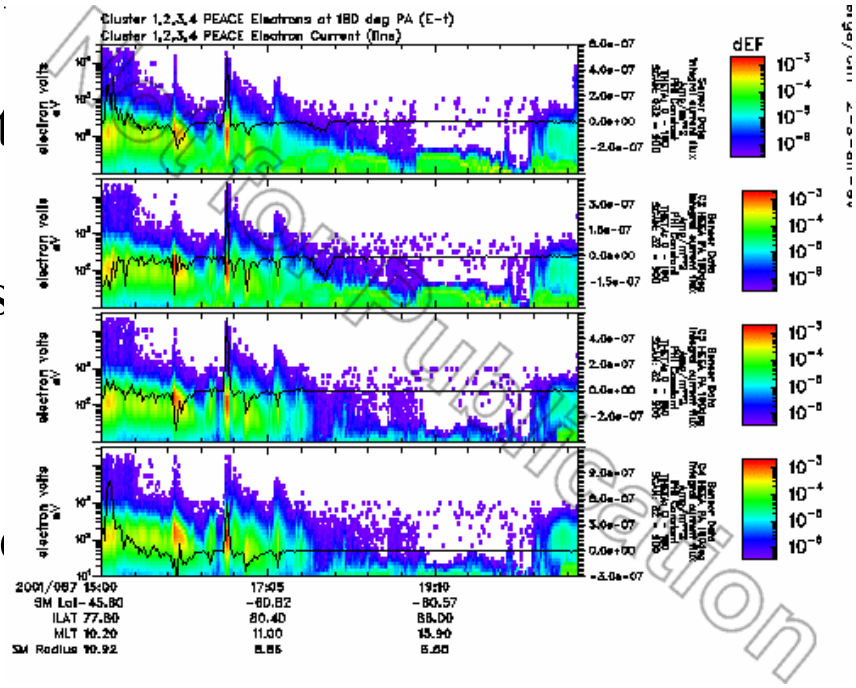
Data Preservation

	Storage Media		Data Safety			
Spacecraft	Short Term	Long Term	Multiple Copies	Multiple Sites	Sites Apart	Archive
SOHO	HD	HD	Y	Y	Y	Y
RHESSI	HD	HD	Y	Y	Y	Y
ACE	HD	CD-R, DVD-R	Y	Y	N	Y
Polar	HD	CD-R, DVD-R	Y	Y	Y	Y
FAST	HD	CD-R, DVD-R	Y	Y	Y	Y
SAMPEX	HD	HD, Opt.	Y	Y	Y	Y
TIMED	HD, Tape	NYD	Y	Y	Y	Y
TRACE	HD	HD	Y	Y	Y	Y
Ulysses	HD, CD-R	HD,CD-R, Tape	Y	Y	Y	Y
IMAGE	HD, CD-R,DVD-R	HD, DVD-R, Tape	Y	Y	Y	Y
Wind	HD	CD-R, DVD-R	Y	Y	Y	Y

HD-Hard Magnetic Disk, **Opt.** –Rewritable Optical Media, **NYD**- Not Yet Decided

Cluster

- Cluster is an international mission lead by ESA.
- Both US co-investigators and data users have found it difficult to obtain data.
- Only browse level data products are available without PI permission.
- It has proven difficult to produce well-calibrated data from the 4 spacecraft.
- ESA has started a program to provide archival quality data.



Where are we?

- Space physics missions largely adhere to the CODMAC principles for successful science data management.
  - In general the data are available to the science community in a timely manner. We are approaching an open data environment.
  - Scientists can find the needed data provided they know where to look.
  - Much more care is being taken to document the data adequately.
  - Care is being taken to make sure that the data are preserved.

Where are we?

- Comparative research within missions isn't as easy as it could be and comparative research between missions is still very difficult.
  - Locating data if you don't know where to go is difficult.
  - Once a researcher locates the required data it comes in multiple formats.
  - There are no commercially or publicly available tools that allow researchers to use all of the formats currently used in space physics.
  - There are no space physics wide metadata standards (the existence of which could significantly lessen the burden of the multiple formats).

How Well Prepared Is Solar Terrestrial Physics to Meet the Demands of Future Missions?

- We are entering an era with unprecedented data management challenges.
  - Much more sophisticated instruments are being proposed.
  - Data volumes are growing exponentially. (Future missions will produce ~10^{15} B of data!)
- Will our current approach to data management work in this new environment?
  - Is mission by mission and instrument by instrument approach to data management optimal ?
  - Are current data system technologies scaleable to planned data rates from future missions?
  - How will access to and preservation of data be maintained after the end of these missions?